Data Science at the University of Virginia—A School Without Walls

Introduction

The School of Data Science (SDS) at the University of Virginia was [established in September 2019](#) with support of the University leadership 2030 [strategic plan](#), and a transformational gift of $120 million—the largest private donation in the university's history. It is driven by a vision to meet the challenges of an increasingly data-driven world through excellence in collaborative, open, interdisciplinary research; education which produces responsible, diverse leaders; and service that leads to products and outcomes which provide societal benefit. The values that guide that vision are excellence, integrity, diversity, openness and transparency, [FAIR](#) data principles, innovation, and dedication to working for the social good.

Defining Data Science

A major part of understanding and appreciating how the school would grow and evolve was coming up with a flexible, yet applicable definition of data science on which to base its research, educational and community programs. Data science is a complex and evolving field, but most agree that it can be defined as the intersection of computer science and technology, math and statistics, and domain knowledge, with the purpose of extracting knowledge and value from data. With this backdrop the school has proposed and is implementing in all we do, a 4+1 model of data science. The "4" are areas that build upon that broad definition, but define the field, at least for us, as it grows and comes to be practiced across disciplines. The areas are defined as **value, design, systems and analytics** with the "+1" the **practice** that connects the four areas. Together, the 4+1 defines specific kinds of expertise that are a part of every data science project but are often unconnected and therefore do not fully benefit from one another in academia given its siloed nature. This major undertaking on the faculty across the team of honing this definition has led to further clarity around educational programs, research directions, overall strategic goals and hiring strategy.

The area of **value** explores the tension between ethics and the professional activities of business planning, policy making, developing motivations for scientific research, and other activities that have a direct impact on people and their environment. This is the area where the school determines what we do versus what we do not do, in order to maximize societal benefit and minimize harm. It is also the area that looks inward to the other data science areas and provides guidance on such issues as algorithmic bias or open science.

The area of **design** includes expertise in human machine interaction as it appears at the points of both consuming data and producing data products. This includes the representation and communication of human reality as data for the work of analytics. It also includes the making of things, with purpose and intent. A key part of this area is the broad practice of what is often called visualization, the translation of complex quantitative information in visual (and other sensory) forms that humans can understand. Beyond visualization, open dissemination and evaluation is part of design.

The area of **systems** includes expertise in infrastructure systems and architectures to support working with big data, in terms of volume, velocity, and variety, and building high performance pipelines in both

development and production environments. It includes the broad areas of hardware, software, networks and cybersecurity from the perspective of practical application rather than theoretical computer science.

The area of **analytics** includes what most consider to be the heart of data science, the combination of statistical methods with machine learning, along with optimization, signal processing, network analysis, and other rigorous quantitative methods from across fields. Although unified by a broad commitment to advanced mathematical methods and algorithms, for us analytics is a collection of complementary methods and goals.

The area of **practice** brings together the four areas and consists of actual applications and activities that bring people together to combine expertise from the other areas. It is characterized by data science teams working together and with external parties, public and private, to develop solutions and projects that are responsible, authentic, effective, and efficient. Practice is also where the core areas of data science encounter domain knowledge and real-world problems. To this end, the school's major focus is in the verticals of environment, biomedicine, democracy (more broadly the digital humanities), business, education and cybersecurity.

Building Upon Institutional Strengths

Beyond our definition of data science is the consideration of how we leverage what the university is known for. This includes the undergraduate experience, our professional schools of law and business, our school of global policy, our work in democracy and a first-rate academic medical center.

Financial Stability

We are working on supporting 4 pillars of academic stability – tuition, direct and indirect costs from grants, public private partnership and philanthropy. It is our belief that if any one of those pillars becomes too dominant in a changing marketplace our sustainability will be in question. We think of it as a balanced portfolio.

Accomplishments

Success of the School of Data Science relies on its collective growth in human capital and collaborative opportunities. The growth in talent has provided opportunities for the development of several new educational programs, including a fully online version of a Master of Science in Data Science to be completed part time for working professionals (complementing our residential MSDS program), an interdisciplinary undergraduate minor in data science, as well as a PhD program in Data Science (currently under state review).

One of the founding principles of the SDS is to disseminate knowledge openly and freely to all, motivated by the desire to maximize its impact. Therefore, the school underwent a process of developing and approving Open Access Guidelines and Recommendations to drive innovation across boundaries and increase access to knowledge. These guidelines were crafted to set expectations for all SDS team members as it relates to the dissemination of their scholarly and academic works, and attribution of such works. We are proud that these guidelines, through approval by the faculty senate, are being adopted across the university. Exactly the kind of impact we wish our model of higher education to achieve.

Not satisfied with being a professional school, an active research program is underway, hiring leaders and faculty with a strong research agenda with emphasis on the "+1". Already we are seeing success as measured by funding and perceived impact, which we hope to measure quantitatively in the future.

Opportunities and Challenges

We find ourselves in a competitive market for the hiring of faculty and research staff to build out in these important areas. Joint hiring between schools and departments is well intentioned, but strategic alignment is an issue and success depends very much on the engagement of the individual faculty member with our school. Established faculty are hard to come by and have high demands and yet are vital to seed our identified areas of data science. This past year we tried to hire 17 new faculty and staff and were successful in hiring 13, losing 2, all at a large service cost on search committees and related activities. This academic year we are aiming at a more target and strategic hiring process versus an open approach.

Building a school is a difficult task which--especially early on--is as much about administration and establishing process as it is academics, and so the elimination or hierarchy, focus on team development, and recognition of the role all people play in the development and working of the school was an essential part of its initial success. The work of building a school that is unlike others within higher education writ large makes it that much more difficult, as each decision is viewed through the lens of traditional academia. Development of policies and procedures, particularly a promotion and tenure policy that values commitment to collaboration, openness and transparency, and translation leads to the development of a new generation of researchers. From there, further development of faculty, especially early career faculty, is important to helping them achieve success according to these new policies. This requires a kind of matrix approach that will allow the diversity of backgrounds to be successful in this endeavor, and hopefully, the broadening of definitions of success, dissemination, and scholarly output across disciplines that utilize data science techniques.

While our approach speaks to equity and inclusion, diversity, particularly amongst our faculty, remains an issue. We took an early step of hiring an Associate Dean for Diversity, Equity and Inclusion as part of a broader initiative. We are not in the business of ticking boxes but moving the needle. Time will tell.

Finally, this past year we undertook strategic planning in the areas of IT and communications led by new hires in these respective areas. This next year will see us revising our overall strategic plan as after what will be two years we see several goals differently – inevitable given fast-moving field and organization. Perhaps also a sign that we are moving from a startup phase to more of a steady state. Time will tell, but whatever else changes we will remain a "School without Walls." That much is in our psyche and something that makes us all very proud.